A Myopic Adjustment Process Leading to Best-Reply Matching

Droste, E.J.R.; Kosfeld, M.; Voorneveld, M.

Publication date:
1998

A Myopic Adjustment Process leading to Best-Reply Matching*

Edward Droste[‡] Michael Kosfeld Mark Voorneveld[§]

Department of Econometrics and CentER, Tilburg University
P.O. Box 90153, 5000 LE Tilburg, The Netherlands

Abstract

We analyze a stochastic process that models myopic strategy adjustment in strategic-form games. It is shown that the steady states of the continuous time limit, which is constructed assuming frequent play and slow adjustment of strategies, are exactly the regret equilibria, as discussed in Droste, Kosfeld, and Voorneveld [4]. This equilibrium concept captures so-called best-reply matching behavior. We derive stability results for the steady states of the continuous time limit in 2×2 games and coordination games. Analyzing the asymptotic behavior of the stochastic adjustment process in discrete time shows convergence to the absorbing states of the adjustment process, which turn out to coincide with the regret equilibria in pure strategies.

Journal of Economic Literature Classification Numbers: C72, D83.

KEYWORDS: Stochastic adjustment process, best reply, matching, regret equilibrium.

*We would like to thank Eric van Damme and Dolf Talman for their useful suggestions.

‡Corresponding author. E-mail: E.J.R.Droste@kub.nl.

§This author's research is financially supported by the Dutch Foundation for Mathematical Research (SWON) through project 613-04-051.

1 Introduction

In this paper we analyze a myopic adjustment process for behavior in repeated interactive situations. The interactive situations are modelled as finite strategic-form games and, at each point in time, the players are characterized by a probability distribution over the set of possible actions. This probability distribution describes the probability with which each of the possible actions is actually chosen. Every time the players interact, each player is confronted with a realized outcome of the game. This outcome typically has a positive or negative value to the player, and is expected to influence, by means of myopic adjustment of the probability distribution over the set of possible actions, the future behavior of that player.

When a player interacts with other players and faces their mixed strategy profiles, he is confronted with uncertainty. Obviously, regret considerations may become important in those situations. We say that a player ends up feeling regret whenever his chosen action ex post turned out not to be a best reply to the realized action profile of his opponents. Therefore, we will assume that a player associates a positive or negative value with an outcome, depending on whether or not his action was a best reply, respectively.

In the course of examining the asymptotic behavior of the adjustment process, which is a Markov process in discrete time, two limits will be taken. First, we let time between two successive interactions approach zero. Assuming that players adjust slowly, i.e., players' adjustments of strategies between two successive interactions are small, we obtain a system of deterministic differential equations, which approximates the original stochastic process in discrete time. We analyze stability properties of this system. Second, we are interested in the asymptotic behavior of the stochastic process itself. We will show that the asymptotic behavior of the system of differential equations and the stochastic process may be quite different. This is similar to Börgers and Sarin [2] who stress the importance of considering these different limits.

A feature of the adjustment process discussed in this paper is that, under certain conditions, it predicts matching behavior in the limit. In fact, it predicts so-called best-reply matching behavior, meaning that each player plays an action with a probability equal to the probability that this action is a best reply to the action profile of his opponents. Matching behavior suggests that a player learns the probabilities with which the action profiles of the opponents occur. However, once these probabilities are 'known', he does not reply by playing the action that maximizes expected utility, but instead matches the probabilities. Consider for example the decision-theoretic setting where a player repeatedly faces a two-armed bandit problem. A two-armed bandit is a slot machine with two arms, each operating with a fixed probability. The

player's task is to predict which arm will be operating. In such a situation matching behavior is supported by experiments. See e.g. Edwards [5], Siegel and Goldstein [16], and Suppes and Atkinson [17]. We propose a model where matching behavior occurs in interactive situations, too.

Our myopic adjustment process is closely related to the literature on *reinforcement learning*. See Bush and Mosteller [3] for an early reference. Reinforcement learning models have received a lot of interest, recently, either to explain experimental findings (Roth and Erev [15], Erev and Roth [6]), or to give learning theoretical foundations of population dynamics used in evolutionary game theory (Börgers and Sarin [1, 2]). The basic idea of these models is that players behave in a *stimulus–response* fashion where actions are positively or negatively *reinforced* through experiences of the player. From the players' rationality or knowledge point of view these models are very simple, since they do not require a player to develop a highly complicated cognitive processes in order to form beliefs about his environment. Instead they rely on cognitively rather limited adaptive behavior. (Erev and Roth [6] call their approach 'low' (rationality) game theory in contrast to standard 'hyper-rational' game theory.) In this sense they coincide with our approach, since we also assume players to myopically adjust their behavior to the experience they make when playing the game. Precisely, we assume that actions are reinforced when they are a best reply to the realized outcome of the game. This adjustment procedure is quite different from other procedures we know of. In particular, cardinal concepts as for example earned payoffs play no role since players focus on best replies only, which is an ordinal concept.

The work of Börgers and Sarin [2] is of special relevance for our model since our techniques are very similar to the ones they use. Moreover, in their co-paper (Börgers and Sarin [1]) the authors obtain a comparable matching result. Still, there approach relies on an endogenously moving aspiration level, whereas our players have a fixed aspiration level based on their best-reply structure.

Another part of the literature to which we see our approach connected comprises the work of Hart and Mas-Colell [10, 11], Fudenberg and Levine [8], and Foster and Vohra [7] on *calibration models*. Yet, while these papers focus mainly on adaptive procedures leading to correlated equilibria, our approach particularly features convergence to best-reply matching as captured by the regret equilibrium concept in Droste, Kosfeld, and Voorneveld [4]. Still, to some part the intuition of our assumptions coincide. For instance Hart and Mas-Colell [10] assume that players adjust probabilities to individual actions by measuring the average regret of not having played that action in the past. Our players are more myopic in the sense that they focus on recent history only. Moreover, regret in Hart and Mas-Colell [10] is measured by payoff differences, whereas our approach measures regret in terms of best-replies. Hart and

Mas-Colell [11] consider also a general class of adaptive procedures.

The paper is organized as follows. In section 2 we introduce the model. Section 3 deals with the case of letting the length of a time period go to zero. The stability properties of the steady states of the resulting continuous time limit are analyzed in section 4. The asymptotic behavior of the discrete-time Markov process is discussed in section 5. Finally, section 6 concludes.

2 Model

We consider a finite strategic game represented by a tuple $G = \langle N, (S_i)_{i\in N}, (\succeq_i)_{i\in N}\rangle$, where $N = \{1, \ldots, n\}$ is a finite set of players. Each player $i \in N$ has a finite set S_i of pure strategies, henceforth actions, and a binary relation $\succeq_i$ over the joint action set $S = \prod_{i\in N} S_i$, reflecting his preferences over the outcomes. The binary relation $\succeq_i$ is assumed to be reflexive and its asymmetric part $\succ_i$, defined for all $s, t \in S$ by

$$s \succ_i t \Leftrightarrow s \succeq_i t \text{ and } t \not\succeq_i s,$$

is assumed to be acyclic. For notational convenience we write $S_{-i} = \prod_{j\in N\setminus\{i\}} S_j$. Given an action tuple $s = (s_1, \ldots, s_n)$ and $i \in N$ we let $s_{-i} = (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n)$ and, with a slight abuse of notation, $s = (s_i, s_{-i})$. By

$$\Delta_i := \{\sigma_i : S_i \to \mathbb{R} \mid \forall s_i \in S_i : \sigma_i(s_i) \geq 0, \sum_{s_i\in S_i} \sigma_i(s_i) = 1\}$$

we denote the set of mixed strategies, henceforth strategies, for player i. Analogous to the action case we use notations Δ, Δ_{-i}, and $\sigma = (\sigma_i, \sigma_{-i})$. By

$$\operatorname{int}(\Delta_i) := \{\sigma_i : S_i \to \mathbb{R} \mid \forall s_i \in S_i : \sigma_i(s_i) > 0, \sum_{s_i\in S_i} \sigma_i(s_i) = 1\}$$

we denote the set of strategies for player i, such that a positive probability is assigned to all actions. For a strategy profile $\sigma_{-i} \in \Delta_{-i}$, we write $\sigma_{-i}(s_{-i}) := \prod_{j\in N\setminus\{i\}} \sigma_j(s_j)$, the probability that the opponents of player i play the action profile $s_{-i} \in S_{-i}$. Finally, denote for each player $i \in N$ and each profile $s_{-i} \in S_{-i}$ of actions of his opponents the set of best replies, i.e. the actions that player i cannot improve upon, by

$$B_i(s_{-i}) := \{s_i \in S_i \mid \nexists \tilde{s}_i \in S_i : (\tilde{s}_i, s_{-i}) \succ_i (s_i, s_{-i})\}.$$

Since S_i is finite and $\succ_i$ is acyclic we know that $B_i(s_{-i})$ is nonempty. We assume that each player $i \in N$ knows the set $B_i(s_{-i})$ for each action profile s_{-i} of his opponents.

The n players play the game G, as introduced above, repeatedly. The iterations of the game are indexed by $k \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$. At each time $k \in \mathbb{N}_0$, each player

$i \in N$ will be characterized by a strategy $\sigma_i^k \in \Delta_i$. We take the players' behavior to be random, meaning that, at each time k, each player i does play an action $s_i^k \in S_i$, but this action is drawn from the probability distribution over his action set induced by his strategy σ_i^k. By $\sigma_i^k(s_i)$ we denote the probability that player i, at time k, plays action s_i. For a strategy profile σ_{-i}^k we write $\sigma_{-i}^k(s_{-i}) := \prod_{j \in N \setminus \{i\}} \sigma_j^k(s_j)$, the probability that the opponents of player i, at time k, play the action profile $s_{-i} \in S_{-i}$. The state of the game $\sigma^k = \left(\sigma_1^k, ..., \sigma_n^k\right) \in \Delta$ at time $k \in \mathbb{N}_0$ identifies a strategy for each player i, with the initial state of the game $\sigma^0 = (\sigma_1^0, ..., \sigma_n^0)$ exogenously given. We assume that, after the kth repetition of the game, each player i observes the realized action profile s_{-i}^k of his opponents.

Now we describe the myopic adjustment process that specifies how σ^k evolves over time. Consider a fixed period $k \in \mathbb{N}_0$ and assume that in this period action profile $s \in S$ was realized. All players $i \in N$ adjust their strategy as follows

$$\sigma_i^{k+1}(s_i) = \begin{cases} (1-\theta)\,\sigma_i^k(s_i) + \frac{\theta}{|B_i(s_{-i})|}, & \text{if } s_i \in B_i(s_{-i}) \\ (1-\theta)\,\sigma_i^k(s_i), & \text{otherwise,} \end{cases} \tag{1}$$

where $0 < \theta < 1$ is exogenously given. For a given initial random variable σ^0 and a given parameter θ, the adjustment rule implies that $\left\{\sigma^k\right\}_{k \in \mathbb{N}_0}$ is a discrete-time Markov process with infinite state space Δ.

The adjustment process specified in (1) states that a player $i \in N$, after the kth repetition of the game, myopically adjusts his strategy by first proportionally decreasing all probabilities by a fraction $0 < \theta < 1$. This then leaves the player with a probability θ that we assume will be reallocated to the actions that are best replies to the action profile of his opponents in the kth repetition of the game. In fact, following the principle of insufficient reason, as first introduced by Jacob Bernoulli (cf. Luce and Raiffa [12], p. 284), we assume that the probability θ will be equally distributed over the best replies. Clearly, the action currently played by player i does not directly influence how he adjusts his strategy, since the adjustment process is completely determined by the current action profile of his opponents. The action currently played by player i does, however, influence the updating process of all of his opponents. Consequently, the action currently played by player i does influence the future strategy profiles of his opponents and therefore it will influence his own adjustment process indirectly. Furthermore, note that an action being a best reply to the action profile of the opponents does not necessarily imply that the probability with which this action is played increases.

For later convenience we end this section by describing the expected movement of the state of the game. Consider the kth repetition of the game and suppose that the current state is $\bar{\sigma}$. Note that, conditional on this, the state in period $k+1$ is

a random variable. Let $E\left[\sigma_i^{k+1}(s_i) - \sigma_i^k(s_i) \mid \sigma^k = \bar{\sigma}\right]$ denote the expected value of $\sigma_i^{k+1}(s_i) - \sigma_i^k(s_i)$ conditional on the state at time k being $\bar{\sigma}$. For all $k \in \mathbb{N}_0$, $\bar{\sigma} \in \Delta$, $i \in N$, and $s_i \in S_i$, it holds that

$$\begin{aligned}
& E\left[\sigma_i^{k+1}(s_i) - \sigma_i^k(s_i) \mid \sigma^k = \bar{\sigma}\right] \\
= & \sum_{\{s_{-i} \in S_{-i} | s_i \in B_i(s_{-i})\}} \left[(1-\theta)\,\bar{\sigma}_i(s_i) + \frac{\theta}{|B_i(s_{-i})|} - \bar{\sigma}_i(s_i)\right] \bar{\sigma}_{-i}(s_{-i}) \\
& + \left[(1-\theta)\,\bar{\sigma}_i(s_i) - \bar{\sigma}_i(s_i)\right] \left[1 - \sum_{\{s_{-i} \in S_{-i} | s_i \in B_i(s_{-i})\}} \bar{\sigma}_{-i}(s_{-i})\right] \\
= & \sum_{\{s_{-i} \in S_{-i} | s_i \in B_i(s_{-i})\}} \left[\frac{\theta}{|B_i(s_{-i})|} - \theta\bar{\sigma}_i(s_i)\right] \bar{\sigma}_{-i}(s_{-i}) \qquad (2) \\
& - \left[\theta\bar{\sigma}_i(s_i)\right] \left[1 - \sum_{\{s_{-i} \in S_i | s_{-i} \in B_i(s_{-i})\}} \bar{\sigma}_{-i}(s_{-i})\right] \\
= & \ \theta \left[\left[\sum_{\{s_{-i} \in S_{-i} | s_i \in B_i(s_{-i})\}} \frac{1}{|B_i(s_{-i})|} \bar{\sigma}_{-i}(s_{-i})\right] - \bar{\sigma}_i(s_i)\right].
\end{aligned}$$

3 Continuous Time Limit

In this section we analyze the behavior of the adjustment process specified in (1), as the length of a time period approaches zero. Taking this limit will allow us to work with a deterministic system in continuous time, which will turn out to be convenient concerning the asymptotic analysis. However, since the deterministic system is an approximation of the original process, it is not clear whether the asymptotic results for the deterministic system are true for the stochastic process as well. In fact, we will show that this is not the case, thereby affirming the difference between both approaches. See also Börgers and Sarin [2]. We assume that the players' adjustments of their strategies slows down at the same rate at which the time difference between the repetitions shrinks. Under this assumption it is appropriate to take this limit since it enables us to ignore the fact that players coordinate their strategy adjustments.

In order to establish the continuous time approximation of the adjustment process specified in (1), we let the time that passes between two repetitions of the game be equal to $\eta > 0$. This gives rise to the following modification of (1) for all $i \in N$ and $k \in \mathbb{N}_0$:

$$\sigma_i^{\eta,k+1}(s_i) = \begin{cases} (1-\eta\theta)\,\sigma_i^{\eta,k}(s_i) + \frac{\eta\theta}{|B_i(s_{-i})|}, & \text{if } s_i \in B_i(s_{-i}) \\ (1-\eta\theta)\,\sigma_i^{\eta,k}(s_i), & \text{otherwise.} \end{cases} \qquad (3)$$

Again, for a given $0 < \theta < 1$, we are left with a discrete time Markov process $\left\{\sigma^{\eta,k}\right\}_{k\in\mathrm{N}_0}$, provided that we specify the initial random variable $\sigma^{\eta,0}$. Note that, since we let the time interval between repetitions of the game be equal to η, the random variable $\sigma^{\eta,k}$ gives the state of the process at time ηk.

We are interested in the limit of (3) as $\eta \to 0$. In fact, analyzing the limit of $\sigma^{\eta,k}$ for any sequence of η's and k's such that $\eta \to 0$ and $k\eta \to t$ gives us the state of the continuous time limit at time $t \in \mathbb{R}_+$. To describe the continuous time limit we introduce a function $\hat{\sigma}^t \in \Delta$, which is differentiable with respect to t. The derivative of $\hat{\sigma}^t$ is given by

$$\frac{d\hat{\sigma}_i^t(s_i)}{dt} = \theta\left[\left[\sum_{\{s_{-i}\in S_{-i}|s_i\in B_i(s_{-i})\}} \frac{1}{|B_i(s_{-i})|}\hat{\sigma}_{-i}^t(s_{-i})\right] - \hat{\sigma}_i^t(s_i)\right], \tag{4}$$

for all $t \in \mathbb{R}_+$, $i \in N$, and $s_i \in S_i$. Let $\hat{\sigma}^0$ denote the initial value of $\hat{\sigma}^t$. The next proposition shows that $\hat{\sigma}^t$ describes the evolution of play under the assumptions made above.

Proposition 1 *Suppose that for all $\eta > 0$ it holds that $\sigma^{\eta,0} = \hat{\sigma}^0$ with probability 1. Consider some t with $0 \leq t < \infty$ and let $\eta \to 0$ and $k\eta \to t$. Then $\sigma^{\eta,k}$ converges in probability to $\hat{\sigma}^t$.*

Proof. It is sufficient to verify that the assumptions of Theorem 1.1 in Chapter 8 of Norman [14] are satisfied. Norman's condition (a.1) is trivially satisfied because the state space of the discrete time Markov process $\left\{\sigma^{\eta,k}\right\}_{k\in\mathrm{N}_0}$ is independent of η. Furthermore, because the functions $v : \Delta \to \mathbb{R}^{\Sigma_{i\in N}|S_i|}$ and $w : \Delta \to \mathbb{R}^{(\Sigma_{i\in N}|S_i|)^N}$, defined by

$$v(\bar{\sigma}) := E\left[\frac{\sigma^{\eta,k+1}-\sigma^{\eta,k}}{\eta}\middle|\sigma^{\eta,k}=\bar{\sigma}\right]$$

and

$$\begin{aligned} w(\bar{\sigma}) &= E\left[\left(\frac{\sigma^{\eta,k+1}-\sigma^{\eta,k}}{\eta}\right)^2\middle|\sigma^{\eta,k}=\bar{\sigma}\right] - E\left[\frac{\sigma^{\eta,k+1}-\sigma^{\eta,k}}{\eta}\middle|\sigma^{\eta,k}=\bar{\sigma}\right]^2 \\ &:= Var\left[\frac{\sigma^{\eta,k+1}-\sigma^{\eta,k}}{\eta}\middle|\sigma^{\eta,k}=\bar{\sigma}\right], \end{aligned}$$

are also independent of η, conditions (a.2) and (a.3) are satisfied, respectively. Note that $Var\left[\frac{\sigma^{\eta,k+1}-\sigma^{\eta,k}}{\eta}\middle|\sigma^{\eta,k}=\bar{\sigma}\right]$ denotes the n-dimensional variance-covariance matrix of the random variable $\frac{1}{\eta}\left(\sigma^{\eta,k+1}-\sigma^{\eta,k}\right)$ conditional on the event that the state of the game in stage k is $\bar{\sigma}$. Norman's conditions (b.1), (b.2), and (b.3), require v to be differentiable, the derivative of v to be bounded, and the derivative of v to be

Lipschitz continuous, respectively. Condition (b.4) states that the function w should also be Lipschitz continuous. Finally, Norman's condition (c) requires the function $r : \Delta \to \mathbb{R}$, defined by

$$r(\bar{\sigma}) = E\left[\left|\frac{\sigma^{\eta,k+1} - \sigma^{\eta,k}}{\eta}\right|^3 \middle| \sigma^{\eta,k} = \bar{\sigma}\right],$$

with

$$|x|^3 := \sum_{l=1}^{\Sigma_{i\in N}|S_i|} |x_l|^3 \text{ if } x \in \mathbb{R}^{\Sigma_{i\in N}|S_i|},$$

to be bounded from above. In our model all these functions are polynomial functions with compact domains, and therefore Norman's conditions (b.1), (b.2), (b.3), (b.4), and (c) are satisfied. ■

The above result states that, if η and $k\eta$ are close to 0 and t, respectively, then, with large probability, $\sigma^{\eta,k}$ will be close to $\hat{\sigma}^t$. So, frequent play and slow movement make it possible to apply a law of large numbers argument, and as a result the actual and expected movement of the adjustment process coincide.

Thus, starting from a simple strategy adjustment procedure that is followed by myopic players in a finite n-person game, we obtain through a well-specified limit taking process a deterministic system that is given by (4).

In order to get a better intuition for the deterministic system we provide an alternative interpretation, which instead of relying on myopic strategy adjustment of a finite number of players relies on a multipopulation model. Suppose there exist n large (infinite) populations of agents. Each agent in population i $(i = 1, ..., n)$ is programmed to an action $s_i \in S_i$. The fraction of agents in population i programmed to action s_i at time t is given by $\hat{\sigma}_i^t(s_i)$. At any instant of time a fraction θ of every population is called to play the game. These agents are randomly matched in n-tuples. After the game has been played each participating agent dies and is replaced by one child. The child adopts a best-reply to the action profile his parent was confronted with. In the case of multiple best replies the child adopts each of them with equal probability. Let $h_i(s_i, \hat{\sigma}_{-i}^t)$ denote the expected value of action s_i in terms of best-replies. That is, fixing the value of an action as 1 if it is a best reply and 0 otherwise,

$$h_i(s_i, \hat{\sigma}_{-i}^t) = \sum_{\{s_{-i}\in S_{-i} | s_i \in B_i(s_{-i})\}} \frac{1}{|B_i(s_{-i})|} \hat{\sigma}_{-i}^t(s_{-i}).$$

With time being continuous the evolution of actions being played in population i is then described by the following differential equations:

$$\frac{\hat{\sigma}_i^t(s_i)}{dt} = \theta h_i(s_i, \hat{\sigma}_{-i}^t) - \theta \hat{\sigma}_i^t(s_i),$$

for all $t \in \mathbb{R}_+$, $i \in N$, and $s_i \in S_i$, which are obviously identical to (4).

The continuous time limit $\hat{\sigma}^t$ is related to the concept of regret equilibrium, as discussed in Droste, Kosfeld, and Voorneveld [4]. This equilibrium notion is defined as follows:

Definition 1 *Let $G = \langle N, (S_i)_{i \in N}, (\succeq_i)_{i \in N} \rangle$ be a game. A mixed strategy profile $\sigma \in \Delta$ is a regret equilibrium if for every player $i \in N$ and for every $s_i \in S_i$:*

$$\sigma_i(s_i) = \sum_{\{s_{-i} \in S_{-i} | s_i \in B_i(s_{-i})\}} \frac{1}{|B_i(s_{-i})|} \sigma_{-i}(s_{-i}). \tag{5}$$

As shown in Droste, Kosfeld, and Voorneveld [4], regret equilibria exist for every game G. The following proposition gives a motivation for considering the concept of regret equilibrium, by stating that the steady states of the continuous time limit are exactly the regret equilibria.

Proposition 2 *Consider a game G. The set of steady states of the continuous time limit $\hat{\sigma}^t$, which is characterized by $\dot{\hat{\sigma}}_i(s_i) := \frac{d\hat{\sigma}_i^t(s_i)}{dt} = 0$ for all $i \in N$ and $s_i \in S_i$, coincides with the set of regret equilibria of G.*

Proof. Proposition 2 follows immediately from comparing the conditions for a steady state of the continuous time limit (4) and the definition of a regret equilibrium (5). ■

4 Stability Analysis

To establish stability properties of single steady states or sets of steady states of the continuous time limit introduced in section 3, we use both the eigenvalue analysis of the Jacobian matrix, evaluated at a steady state, and the direct Lyapunov method. The eigenvalue analysis is based on the well-known result that a steady state of a system of differential equations is (asymptotically) stable if the real parts of the eigenvalues of the Jacobian matrix, evaluated at the steady state, are all (strictly) negative. Furthermore, a steady state is unstable if there exists an eigenvalue of the Jacobian matrix, evaluated at the steady state, with a strictly positive real part. With respect to the direct Lyapunov method we use the following result.

Theorem 3 [Theorem 6.4 in Chapter 6 of Weibull [18]] *Suppose that $A \subset \Delta$ is closed. If there exists a neighborhood D of A and a continuously differentiable function $v : D \to \mathbb{R}_+$ such that*

$$v(\hat{\sigma}) = 0 \text{ if and only if } \hat{\sigma} \in A$$

and

$$\sum_{i\in N}\sum_{s_i\in S_i}\frac{dv\left(\hat{\sigma}\right)}{d\hat{\sigma}_i\left(s_i\right)}\dot{\hat{\sigma}}_i\left(s_i\right)<0 \text{ for all } \hat{\sigma}\notin A,$$

then A is asymptotically stable. Suppose that $A\subset\Delta$ is closed and connected. If there exists a neighborhood D of A and a continuously differentiable function $v:D\to\mathbb{R}_+$ such that

$$v\left(\hat{\sigma}\right)=0 \text{ if and only if } \hat{\sigma}\in A$$

and

$$\sum_{i\in N}\sum_{s_i\in S_i}\frac{dv\left(\hat{\sigma}\right)}{d\hat{\sigma}_i\left(s_i\right)}\dot{\hat{\sigma}}_i\left(s_i\right)\leq 0 \text{ for all } \hat{\sigma}\notin A,$$

then A is Lyapunov stable.

In general it is not possible to derive explicit results with respect to the stability properties of the steady states of the continuous time limit. Hence, we consider two specific classes of games, namely 2×2 games and coordination games.

4.1 2×2 Games

A 2×2 game is a two-player strategic-form game where the action set of both players consist of two elements. Denoting $S_1=\{T,B\}$ and $S_2=\{L,R\}$ the continuous time limit, characterized by (4), corresponding to a 2×2 game becomes

$$\begin{aligned}\dot{\hat{\sigma}}_1\left(T\right) &= \theta\left[\left[\sum_{\{s_2\in S_2|T\in B_1(s_2)\}}\frac{1}{|B_1(s_2)|}\hat{\sigma}_2^t\left(s_2\right)\right]-\hat{\sigma}_1^t\left(T\right)\right] \\ \dot{\hat{\sigma}}_2\left(L\right) &= \theta\left[\left[\sum_{\{s_1\in S_1|L\in B_2(s_1)\}}\frac{1}{|B_2(s_1)|}\hat{\sigma}_1^t\left(s_1\right)\right]-\hat{\sigma}_2^t\left(L\right)\right],\end{aligned} \tag{6}$$

for all $t\in\mathbb{R}_+$. Note that the equations for $\dot{\hat{\sigma}}_1\left(B\right)$ and $\dot{\hat{\sigma}}_2\left(R\right)$ are redundant since $\hat{\sigma}_1^t\left(B\right):=1-\hat{\sigma}_1^t\left(T\right)$ and $\hat{\sigma}_2^t\left(R\right):=1-\hat{\sigma}_2^t\left(L\right)$. Let σ^* denote a steady state of the continuous time limit (6) or, equivalently, a regret equilibrium of the associated 2×2 game. The following result states the stability properties of the steady states of the continuous time limit of a 2×2 game.

Proposition 4 *Let G be a 2×2 game. Then a steady state σ^* of the continuous time limit (6) is Lyapunov stable.*

Proof. To prove Proposition 4 we use the eigenvalue analysis. The Jacobian matrix $\mathbf{J}_{\hat{\sigma}^t}\left(\sigma^*\right)$, evaluated in a steady state σ^*, is given by

$$\mathbf{J}_{\hat{\sigma}^t}\left(\sigma^*\right)=\left.\begin{bmatrix}\frac{d\dot{\hat{\sigma}}_1(T)}{d\hat{\sigma}_1^t(T)} & \frac{d\dot{\hat{\sigma}}_1(T)}{d\hat{\sigma}_2^t(L)}\\ \frac{d\dot{\hat{\sigma}}_2(L)}{d\hat{\sigma}_1^t(T)} & \frac{d\dot{\hat{\sigma}}_2(L)}{d\hat{\sigma}_2^t(L)}\end{bmatrix}\right|_{\hat{\sigma}^t=\sigma^*}=\begin{bmatrix}-\theta & \left.\frac{d\dot{\hat{\sigma}}_1(T)}{d\hat{\sigma}_2^t(L)}\right|_{\hat{\sigma}^t=\sigma^*}\\ \left.\frac{d\dot{\hat{\sigma}}_2(L)}{d\hat{\sigma}_1^t(T)}\right|_{\hat{\sigma}^t=\sigma^*} & -\theta\end{bmatrix}.$$

Consider the entry $\left.\frac{d\dot{\hat{\sigma}}_1(T)}{d\hat{\sigma}_2^t(L)}\right|_{\hat{\sigma}^t=\sigma^*}$. Depending on the best-reply structure of the game this expression equals $-\theta$, $-\frac{\theta}{2}$, 0, $\frac{\theta}{2}$, or θ. To illustrate why this holds consider the case where T is the unique best reply to L and both T and B are best replies to R. For this best-reply structure it holds that the first equation of the continuous time limit (6) equals

$$\begin{aligned}\dot{\hat{\sigma}}_1(T) &= \theta\left[\hat{\sigma}_2^t(L)+\frac{1}{2}\hat{\sigma}_2^t(R)-\hat{\sigma}_1^t(T)\right]\\ &= \theta\left[\hat{\sigma}_2^t(L)+\frac{1}{2}\left(1-\hat{\sigma}_2^t(L)\right)-\hat{\sigma}_1^t(T)\right].\end{aligned}$$

It follows immediately that in this case $\left.\frac{d\dot{\hat{\sigma}}_1(T)}{d\hat{\sigma}_2^t(L)}\right|_{\hat{\sigma}^t=\sigma^*}=\frac{\theta}{2}$. Considering the other eight possible best-reply structures gives the required result. Similarly, it can be shown that the entry $\left.\frac{d\dot{\hat{\sigma}}_2(L)}{d\hat{\sigma}_1^t(T)}\right|_{\hat{\sigma}^t=\sigma^*}$ also equals $-\theta$, $-\frac{\theta}{2}$, 0, $\frac{\theta}{2}$, or θ.

We are left with the actual eigenvalue analysis of the Jacobian matrix. In determining the eigenvalues we distinguish five case. First, let $\left.\frac{d\dot{\hat{\sigma}}_1(T)}{d\hat{\sigma}_2^t(L)}\right|_{\hat{\sigma}^t=\sigma^*}$ or $\left.\frac{d\dot{\hat{\sigma}}_2(L)}{d\hat{\sigma}_1^t(T)}\right|_{\hat{\sigma}^t=\sigma^*}$ be equal to 0. This implies that the Jacobian matrix is triangular and the eigenvalues are given by the entries on the diagonal, i.e., $\lambda_{1,2}=-\theta<0$. Consequently, a steady state is asymptotically stable.

Second, let $\left.\frac{d\dot{\hat{\sigma}}_1(T)}{d\hat{\sigma}_2^t(L)}\right|_{\hat{\sigma}^t=\sigma^*}=-\theta$. In this case the eigenvalues of the Jacobian matrix are given by

$$\lambda_{1,2}=-\theta\pm\sqrt{-\theta\left.\frac{d\dot{\hat{\sigma}}_2(L)}{d\hat{\sigma}_1^t(T)}\right|_{\hat{\sigma}^t=\sigma^*}}.$$

When $\left.\frac{d\dot{\hat{\sigma}}_2(L)}{d\hat{\sigma}_1^t(T)}\right|_{\hat{\sigma}^t=\sigma^*}=-\theta$ we find that $\lambda_1=-2\theta$ and $\lambda_2=0$, which implies a steady state to be Lyapunov stable, but not asymptotically stable. In all other cases, i.e., $\left.\frac{d\dot{\hat{\sigma}}_2(L)}{d\hat{\sigma}_1^t(T)}\right|_{\hat{\sigma}^t=\sigma^*}$ equal to $-\frac{\theta}{2}$, $\frac{\theta}{2}$, or θ, it holds that $\mathrm{Re}(\lambda_{1,2})<0$, implying a steady state to be asymptotically stable.

Third, let $\left.\frac{d\dot{\hat{\sigma}}_1(T)}{d\hat{\sigma}_2^t(L)}\right|_{\hat{\sigma}^t=\sigma^*}=-\frac{\theta}{2}$. A straightforward calculation shows that the resulting eigenvalues equal

$$\lambda_{1,2}=-\theta\pm\sqrt{-\frac{\theta}{2}\left.\frac{d\dot{\hat{\sigma}}_2(L)}{d\hat{\sigma}_1^t(T)}\right|_{\hat{\sigma}^t=\sigma^*}}.$$

Furthermore, it follows immediately that for all possible values of $\left.\frac{d\dot{\hat{\sigma}}_2(L)}{d\hat{\sigma}_1^t(T)}\right|_{\hat{\sigma}^t=\sigma^*}$, it holds that $\mathrm{Re}(\lambda_{1,2})<0$, i.e., a steady state is asymptotically stable.

Fourth, let $\left.\frac{d\dot{\hat{\sigma}}_1(T)}{d\hat{\sigma}_2^t(L)}\right|_{\hat{\sigma}^t=\sigma^*}=\frac{\theta}{2}$. In this case the eigenvalues are given by

$$\lambda_{1,2}=-\theta\pm\sqrt{\frac{\theta}{2}\left.\frac{d\dot{\hat{\sigma}}_2(L)}{d\hat{\sigma}_1^t(T)}\right|_{\hat{\sigma}^t=\sigma^*}}.$$

Again, this implies that $\mathrm{Re}(\lambda_{1,2}) < 0$ for all possible values of $\left.\frac{d\dot{\hat{\sigma}}_2(L)}{d\hat{\sigma}_1^t(T)}\right|_{\hat{\sigma}^t=\sigma^*}$. Consequently, a steady state is asymptotically stable.

Finally, let $\left.\frac{d\dot{\hat{\sigma}}_1(T)}{d\hat{\sigma}_2^t(L)}\right|_{\hat{\sigma}^t=\sigma^*} = \theta$. This gives rise to the following eigenvalues

$$\lambda_{1,2} = -\theta \pm \sqrt{\theta \left.\frac{d\dot{\hat{\sigma}}_2(L)}{d\hat{\sigma}_1^t(T)}\right|_{\hat{\sigma}^t=\sigma^*}}.$$

From the above expression it can easily be deduced that $\mathrm{Re}(\lambda_{1,2}) < 0$, so that a steady state is asymptotically stable, except when $\left.\frac{d\dot{\hat{\sigma}}_2(L)}{d\hat{\sigma}_1^t(T)}\right|_{\hat{\sigma}^t=\sigma^*} = \theta$. In the latter case we find that $\lambda_1 = -2\theta$ and $\lambda_2 = 0$, which implies a steady state to be Lyapunov stable, but not asymptotically stable. ■

4.2 Coordination Games

A two-player game is a coordination game if both players i, $i \in N = \{1, 2\}$, have the same set of actions, $S_1 = S_2 =: X$, and the unique best reply to an action of the opponent is to play the same action. In the case of a coordination game the continuous time limit (4) is given by

$$\dot{\hat{\sigma}}_i(s_i) = \theta\left(\hat{\sigma}_j^t(s_i) - \hat{\sigma}_i^t(s_i)\right), \tag{7}$$

for all $t \in \mathbb{R}_+$, $i, j \in N$, $j \neq i$ and $s_i \in X$. First, consider the set of all steady states of the continuous time limit (7), i.e., $A = \{\hat{\sigma} \in \Delta \mid \hat{\sigma}_1(s_i) = \hat{\sigma}_2(s_i) \text{ for all } s_i \in X\}$. It is easily seen that the set A is closed. Let the Lyapunov function $v : \Delta \to \mathbb{R}_+$ be given by

$$v(\hat{\sigma}) = \frac{1}{2}\sum_{s_i \in X}\left(\hat{\sigma}_2(s_i) - \hat{\sigma}_1(s_i)\right)^2.$$

Obviously, $v(\hat{\sigma})$ is continuously differentiable and $v(\hat{\sigma}) = 0$ if and only if $\hat{\sigma} \in A$. Furthermore, it holds that

$$\sum_{i \in N}\sum_{s_i \in X}\frac{dv(\hat{\sigma})}{d\hat{\sigma}_i(s_i)}\dot{\hat{\sigma}}_i(s_i) = -2\theta\sum_{s_i \in X}\left(\hat{\sigma}_2(s_i) - \hat{\sigma}_1(s_i)\right)^2 < 0,$$

for all $\hat{\sigma} \notin A$. According to Theorem 3 the set A is asymptotically stable. In fact, since $D = \Delta$ the set A is even globally asymptotically stable.

Second, let A be the closed and connected set consisting of a single regret equilibrium, i.e., $A = \{(\bar{\sigma}_1, \bar{\sigma}_1)\}$. Let the Lyapunov function $v : \Delta \to \mathbb{R}_+$ be given by

$$v(\hat{\sigma}) = \frac{1}{2}\sum_{s_i \in X}\left(\hat{\sigma}_1(s_i) - \bar{\sigma}_1(s_i)\right)^2 + \left(\hat{\sigma}_2(s_i) - \bar{\sigma}_1(s_i)\right)^2.$$

Again, it is obvious that $v(\hat{\sigma})$ is continuously differentiable, $v(\hat{\sigma}) = 0$ if and only if $\hat{\sigma} \in A$, and

$$\sum_{i \in N} \sum_{s_i \in X} \frac{dv(\hat{\sigma})}{d\hat{\sigma}_i(s_i)} \dot{\hat{\sigma}}_i(s_i) = -\theta \sum_{s_i \in X} \left(\hat{\sigma}_2(s_i) - \hat{\sigma}_1(s_i)\right)^2 \leq 0,$$

for all $\hat{\sigma} \notin A$. Consequently, according to Theorem 3 each single steady state of (7) is globally Lyapunov stable.

5 Asymptotic Analysis

This section deals with the asymptotic behavior of the stochastic adjustment process specified in (3) as $k \to \infty$ for a fixed η. W.l.o.g. we take $\eta = 1$, hence the stochastic process is given by (1). Taking this limit means that we focus on the stationary distributions of the Markov process in discrete time. In fact, we consider the absorbing states of the stochastic process. These special kind of stationary distributions are states that, once entered, cannot be left.

Proposition 5 identifies the absorbing states of the process (1) as being the regret equilibria in actions of the game G. It follows from Proposition 5.1 in Droste, Kosfeld, and Voorneveld [4] that the regret equilibria in actions are exactly the strict Nash equilibria introduced by Harsanyi [9]. Strict Nash equilibria are those strategy profiles σ satisfying the condition that each player plays his unique best reply to the strategies of his opponents, i.e.,

$$\forall i \in N : \{\sigma_i\} = \{\tau_i \in \Delta_i \mid \nexists \tilde{\tau}_i \in \Delta_i : (\tilde{\tau}_i, \sigma_{-i}) \succ_i (\tau_i, \sigma_{-i})\}.$$

It is clear that strict Nash equilibria, or equivalently, regret equilibria in actions, do not always exist.

Proposition 5 *Consider a game G. If G has at least one strict Nash equilibrium in actions then for every $0 < \theta < 1$ and for any initial state $\sigma^0 \in int(\Delta)$ the stochastic process specified in (1) converges with probability 1 to a strict Nash equilibrium. Furthermore, for any initial state $\sigma^0 \in int(\Delta)$ each strict Nash equilibrium has a positive probability of being the limit of (1).*

Proof. It can easily be verified that the set of regret equilibria in actions of the game G is exactly the set of absorbing states of the stochastic process specified in (1). Theorem 2.3 in Norman [13] says that under certain conditions the stochastic process will converge with probability 1 to one of its absorbing states. Thus, in order to prove the first part of the proposition we are left with verifying the conditions of Norman's theorem.

Condition (H1) is trivially satisfied by the definition of the stochastic process (1). Condition (H2) is satisfied because S_i is finite for all $i \in N$. Condition (H3) requires the stochastic process specified in (1) to be memory-less and temporally homogeneous, meaning that the probabilities of all possible action profiles at time k depend only on the state of the game at time k, i.e. σ^k, and not on earlier states or action profiles, or on the number of repetitions of the game. Since the stochastic adjustment process defined by (1) is a Markov process, condition (H3) is clearly satisfied. Conditions (H4) and (H5) are satisfied because Δ_i is a compact subset of $\mathbb{R}^{|S_i|-1}$ for all $i \in N$. As the function $\varphi_s : \Delta \to \mathbb{R}$ defined by $\varphi_s\left(\sigma^k\right) := \prod_{i=1}^{N} \sigma_i^k\left(s_i\right)$ is differentiable on Δ, condition (H6) only requires the absolute value of each component of $\frac{d\varphi_s\left(\sigma^k\right)}{d\sigma^k}$ to be finite. In fact, it can easily be verified that the absolute value of each component of $\frac{d\varphi_s\left(\sigma^k\right)}{d\sigma^k}$ is an element of $[0,1]$.

Norman's condition (H7) requires the following. Consider any time k and let σ^k and $\tilde{\sigma}^k$ be two possible states of the game at that time. Consider also a fixed action profile $s \in S$. Denote by σ^{k+1} and $\tilde{\sigma}^{k+1}$ the states of the game that are reached if the preceding states were σ^k and $\tilde{\sigma}^k$, respectively, and action profile s was realized at time k. Norman's theorem requires that

$$\left\|\sigma^{k+1} - \tilde{\sigma}^{k+1}\right\| \leq \left\|\sigma^k - \tilde{\sigma}^k\right\|,$$

where $\|\cdot\|$ denotes the Euclidean norm. A straightforward calculation shows that for our adjustment process it holds that

$$\left\|\sigma^{k+1} - \tilde{\sigma}^{k+1}\right\| = (1-\theta)^2 \left\|\sigma^k - \tilde{\sigma}^k\right\|.$$

Since in our model $0 < \theta < 1$, condition (H7) is satisfied. Condition (H8) requires the above weak inequality to be strict in certain cases. Because in our model the inequality is always strict, condition (H8) is also satisfied.

Norman's condition (H9) is not relevant for Theorem 2.3. Finally, condition (H10) states the following. For any initial state $\sigma^0 \in \text{int}(\Delta)$ the closure of the set of states that can be reached from σ^0 with positive probability within finite time, contains at least one of the absorbing states. This condition is satisfied because for a completely mixed initial state the probability that any action profile is realized m times is positive for all $m \in \mathbb{N}$. Consequently, this also holds for an action profile corresponding to an absorbing state. Because realizing the action profile corresponding to an absorbing state any finite number of times generates a sequence of states converging to that absorbing state, condition (H10) is satisfied. This proves the first part of the proposition.

Following the analysis in section 7.2 of Bush and Mosteller [3] it follows immediately that for all initial variables $\sigma^k \in \text{int}(\Delta)$ each absorbing state has a positive

probability of being the limit of the stochastic process (1). This proves the second part of the proposition. ■

Proposition 5 indicates that the asymptotic behavior of the stochastic adjustment process specified in (1) may be different from the asymptotic behavior of the continuous time limit $\hat{\sigma}^t$. To illustrate this point, consider a coordination game as discussed in section 4.2. Due to the stability properties of both each single steady state and the set of steady states of the continuous time limit, a strategy profile close to the steady state $\left(\left(\frac{1}{2},\frac{1}{2}\right),\left(\frac{1}{2},\frac{1}{2}\right)\right)$ will converge to a regret equilibrium in the interior of the state space under the continuous time limit. According to Proposition 5, however, the same strategy profile will converge to one of the two regret equilibria in actions of the coordination game under the stochastic process in discrete time.

6 Concluding Remarks

We have analyzed a stochastic adjustment model in which both taking the limit $\eta \to 0$ and taking the limit $k \to \infty$ is important. Taking the limit $\eta \to 0$ and assuming slow movement results in a continuous time limit. The steady states of the continuous time limit turn out to be exactly the regret equilibria of the corresponding game. In addition, stability results with respect to the steady states of the continuous time limit were derived for 2×2 games and coordination games. Therefore, the adjustment process predicts best-reply matching behavior in the limit and thus gives a motivation for considering the notion of regret equilibrium.

Taking the limit $k \to \infty$ while assuming existence of absorbing states, we show that the process actually converges to one of them. To which one the process will actually converge, however, is not clear. If the initial strategy profile is completely mixed, each absorbing state will be the limit with positive probability. Finally, the absorbing states of the adjustment process coincide with the regret equilibria in actions, or equivalently, the strict Nash equilibria, of the associated game.

References

[1] Börgers, T., and Sarin, R. (1995). "Naive Reinforcement learning with Endogenous Aspirations," mimeo, Department of Economics, University College London.

[2] Börgers, T., and Sarin, R. (1997). "Learning Through Reinforcement and Replicator Dynamics," *Journal of Economic Theory* **77**, 1-14.

[3] Bush, R.R., and Mosteller, F. (1955). *Stochastic Models for Learning*. New York: Wiley.

[4] Droste, E., Kosfeld, M., and Voorneveld, M. (1998). "Regret Equilibria in Games," CentER Discussion Paper 9819, Tilburg University.

[5] Edwards, W. (1956). "Reward Probability, Amount, and Information as Determiners of Sequential Two-Alternative Decision," *Journal of Experimental Psychology* **52**, 177-188.

[6] Erev, I., and Roth, A.E. (1998). "Predicting how People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," *American Economic Review*, forthcoming.

[7] Foster, D.P., and Vohra, R.V. (1997). "Calibrated Learning and Correlated Equilibrium," *Games and Economic Behavior* **21**, 40-55.

[8] Fudenberg, D., and Levine, D. (1996). "Conditional Universal Consistency," *Games and Economic Behavior*, forthcoming.

[9] Harsanyi, J.C. (1973). "Games with Randomly Distributed Payoffs: A New Rationale for Mixed Strategy Equilibrium Point," *International Journal of Game Theory* **2**, 1-23.

[10] Hart, S., and Mas-Colell, A. (1997). "A Simple Adaptive Procedure leading to Correlated Equilibrium," Discussion Paper 126, Hebrew University Jerusalem.

[11] Hart, S., and Mas-Colell, A. (1998). "A General Class of Series leading to Correlated Equilibrium," Session given at the 3rd Spanish Meeting on Game Theory and Applications, Barcelona, June 15-17, 1998.

[12] Luce, R.D., and Raiffa, H. (1957). *Games and Decisions*. New York: Wiley.

[13] Norman, M.F. (1968). "Some Convergence Theorems for Stochastic Learning Models with Distance Diminishing Operators," *Journal of Mathematical Psychology* **5**, 61-101.

[14] Norman, M.F. (1972). *Markov Processes and Learning Models*. New York: Academic Press.

[15] Roth, A., and Erev, I. (1995). "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term," *Games and Economic Behavior* **8**, 164-212.

[16] Siegel, S., and Goldstein, D.A. (1959). "Decision Making Behaviour in a Two-Choice Uncertain Outcome Situation," *Journal of Experimental Psychology* **57**, 37-42.

[17] Suppes, P., and Atkinson, R. (1960). *Markov Learning Models for Multiperson Interaction*. Stanford: Stanford University Press.

[18] Weibull, J.W. (1995). *Evolutionary Game Theory*. Cambridge, MA: MIT Press.